

Taking an “Intentional Stance” on Moral Psychology

Jordan Theriault and Liane Young

A key question of interest to philosophers, and more recently psychologists and neuroscientists, is how we go about attributing complex mental states to the entities (including people) we encounter. Typically, unless we are told explicitly what a person thinks, believes, or desires, our access to information is limited to the observation of behavior. For example, when a man reaches for his pen, we see an arm extended from a body, and fingers wrapped around the pen. Yet we also “see” beyond the surface properties of the action to internal psychological states; we might infer that the man wanted the pen and maybe even for a particular purpose. We might make further behavioral predictions as well (e.g., the man planned to write something down) even in the absence of direct physical evidence (e.g., a piece of paper). By making inferences about mental states that extend beyond what is directly observable, we can be said to have a “theory of mind”—we make inferences about people’s internal mental lives (Baron-Cohen 2001; Onishi and Baillargeon 2005; Saxe 2009).

Our focus in this chapter is not to explain theory of mind—how we represent others’ specific mental states. Instead we want to focus on the broad categories of how we represent other minds. That is, we want to focus not on what people believe or feel specifically but on whether people have minds that render them capable of believing or feeling. An important means by which we represent other minds is by adopting what Dennett (1987) calls an *intentional stance*. To take an intentional stance is to treat an entity *as though it has a mind*. In taking an intentional stance, we effectively change our approach to dealing with the entity in order to more accurately explain and predict its

behavior. Adopting an intentional stance might be automatic in many cases, driven by low-level cues (e.g., contingent movement, the presence of eyes) that signal goal-directed behavior (Arico et al. 2011; Hamlin et al. 2007; Heider and Simmel 1944; Fiala et al. 2011), but the deployment of an intentional stance can also be consciously motivated, in the absence of external cues, to assist in the prediction of behavior. When an entity behaves in a sufficiently complex manner, the abstraction of an intentional stance allows us to simplify the problem into the following: what would a rational agent do? (Dennett 1987). Importantly, this definition does not require us to apply the intentional stance only to entities that are actually capable of possessing minds. Thus, we can think of a clock or a car as having “a mind of its own” or “its own agenda” even while we understand that certain physical properties might be necessary for mental states to really exist (e.g., a brain).

While an intentional stance suggests that we can treat entities as either having a mind or not, in a dichotomous fashion, recent work in philosophy and psychology has proposed that our representation of minds is not sufficiently captured by one dimension. In fact, our attribution of mental states may be best explained along two dimensions: *Agency* and *Experience* (Gray et al. 2007; Gray et al. 2011; Gray et al. 2012a; Gray et al. 2012b; Gray and Wegner 2011a; Knobe and Prinz 2008; Robbins and Jack 2006; Jack and Robbins 2012). Agency is described as the capacity for purposeful action, and goal-directed behavior (roughly analogous to the target of an intentional stance), while Experience is described as the capacity for sensations and feelings, such as pain and pleasure (Gray et al. 2007; Gray et al. 2011; Gray and Wegner 2011a). For example, we may perceive a robot as capable of forming complex goals, but as less capable of experiencing pain, fear, or other sensations (Gray et al. 2007). By contrast, we might easily attribute sensations to a human baby, even while we recognize its limitations in forming and acting on intentions. These two dimensions of mind attribution—Agency and Experience—may clarify our understanding not just of folk psychology, including our inferences about intentional action, but of folk morality as well (Gray et al. 2012a; Gray et al. 2012b).

In this chapter, we will review the literature on mental state attribution from a psychological perspective. In particular, we will examine the intentional stance (Dennett 1987), where we treat an entity as functionally having a mind, and

expect that it will work toward achieving its goals. We will then explore why an intentional stance alone is insufficient to explain how we attribute minds to others, particularly, why the intentional stance fails to capture our sense of moral concern for others. Moral concern is more completely accounted for by the addition of a phenomenological stance, which, put briefly, is the attribution of emotional experience to others (Robbins and Jack 2006). Paralleling the distinction between an intentional and phenomenological stance are the dimensions of Agency and Experience (Gray et al. 2007; Gray et al. 2011; Gray et al. 2012a; Gray et al. 2012b), which pair moral responsibility with the attribution of Agency, and moral rights with the attribution of Experience. Finally, we will review recent work by Sytsma and Machery (2012), which argues that both Agency and Experience are essential to attributing moral standing (i.e. granting moral rights). This stands in contrast to prior work, which has placed moral standing almost exclusively under the domain of attributions of Experience (Gray et al. 2007; Robbins and Jack 2006).

1 The intentional stance

Taking an intentional stance (Dennett 1987) means predicting and explaining observed behavior in terms of what we know about minds and intentional actors. This approach stands in contrast to cases where we might take a *physical stance* or a *design stance*. In the case of a physical stance, we make predictions based on what we know to be true about physical properties (either through folk physics, or scientific knowledge). Taking a physical stance, if we see one billiard ball roll toward another, we know the first ball will impart force to the second ball and move it. Taking a design stance means making predictions based on the typical functioning of a particular entity, without necessary reference to the entity’s underlying physics. If we set an alarm clock to ring in the morning, we predict it will ring at the set time, even without an understanding of the alarm clock’s physical operation. The alarm clock has a clear design, acts predictably, and can be dealt with in a way that is independent of its underlying physical properties. Only when our alarm clock breaks down or acts unpredictably do we consider taking it apart to attempt to understand its hardware (Dennett 1981a).¹

Minds are far too complex for a physical stance or a design stance to make reasonable predictions. Thus, we can adopt an intentional stance, and make predictions based on what we know to be true about minds. Chief among this knowledge is that intentional agents have goals and will work to achieve those goals. Thus, we can expect that agents for whom we adopt an intentional stance will act in whatever way is most likely to achieve their goals. Dennett (1981a) provides the example of a chess-playing computer; to play the computer, we must adopt an intentional stance. Understanding of the physics underlying the operation of the computer—the physical stance—will not be helpful to the player. And to a novice player the means by which the computer selects moves is opaque to the point that he could not predict actions based on his knowledge of how the computer is supposed to work—the design stance. In order to play against the computer, the player adopts an intentional stance, attributing intentions and goals to the computer. The player can then in turn devise strategies that take these intentions into account. Importantly, players adopt an intentional stance despite the fact that the computer lacks the cognitive architecture that we would typically associate with the capacity to represent thought (i.e. a brain). In fact, whether the computer can actually think is irrelevant to the adoption of an intentional stance; from the perspective of the player, treating the computer as though it has a mind is the only means by which he can make predictions about how the computer will behave.

Our utilization of an intentional stance to understand complex behaviors does not necessarily mean that it is appropriate to do so in all cases, or that it will produce the best outcomes. There are certainly cases in which we adopt an intentional stance to our own detriment. The 1997 chess tournament between Garry Kasparov, the world chess champion, and Deep Blue, a chess playing computer designed by IBM (as described by Silver 2012) provides an illustrative example of the potential disadvantages of adopting an intentional stance. In a previous tournament against Deep Blue, Kasparov had consistently taken advantage of his knowledge of how the computer operated: Deep Blue would base its early game strategy on archived data of all previous tournament chess matches that had been played. By playing opening moves that were rarely used in tournaments, Kasparov was able to quickly put the computer in unfamiliar situations. As such Kasparov was making good use of a design stance, where

the actions of the computer could in fact be predicted based on knowledge about how it operates and makes decisions.

Despite this design stance knowledge, Kasparov was allegedly thrown at the end of the first 1997 match, after he had secured an advantageous position and was likely (but not guaranteed) to secure a victory (Silver 2012). Rather than playing to a potential draw, Deep Blue made the unusual and suicidal move of driving a rook into Kasparov's line. Deep Blue forfeited the game after only a few more turns. According to Silver (2012), the strangeness of the move shook Kasparov, as this massive error in a simple position could not be explained by how Kasparov understood the computer to operate. The possible explanations for this bizarre behavior were either that the computer was hiding its capabilities intentionally, that the programmers had thrown the game to make Kasparov overconfident, or else the computer was massively more powerful than previously imagined, and could see ahead so many moves that it had found an alternate route to victory. Either the computer was acting beyond its design as previously understood (by playing mind games), or it was so massively powerful in predicting moves that a simple understanding of its design could no longer suffice (e.g., if it could see more than 20 moves into the future). In the next match, Kasparov famously accused IBM of secretly allowing a grandmaster to make moves on behalf of the machine. Arguably, when the machine failed to conform to the predictions of a design stance, Kasparov adopted an intentional stance toward Deep Blue. In a stroke of irony, the suicidal movement of the rook by Deep Blue was actually a bug: a randomly selected move, which was the result of Deep Blue's indecision. But the damage was done: Kasparov was moved to an intentional stance, and without his knowledge the IBM engineers fixed the bug before the match the next morning (Silver 2012).

That we deploy an intentional stance to predict behavior (or at least attempt it) is further supported by recent work in psychology (Waytz et al. 2010). Waytz and colleagues ran a series of studies in which participants made judgments of the mental capacities of nonhumanoid robots. In their fifth study, the experimenters showed participants a series of videos where robots performed an action, and either asked participants to predict the actions of the robots (with bonus payments for correct answers), or simply asked participants to watch the videos. All participants were shown an initial segment of the video, where the robot performed part of an action (e.g. cleared dishes from the table),

after which the video was paused, and two potential outcomes were presented, (e.g. the robot will either put the dishes in the drawers, or put them on the counter). Participants selected an outcome (or simply read the outcomes in the control condition), and then watched the remainder of the video. Participants then rated the robot on several anthropomorphizing dimensions, including the degree to which the robot had a mind of its own, had consciousness, and possessed intentions, desires, and emotions. Participants who were paid to predict the actions of the robots anthropomorphized them significantly more than the control group. Regardless of whether participants made a conscious decision to take an intentional stance, taking an intentional stance appears to be the consequence of being motivated to understand an entity.

In contrast to Waytz et al. (2010), where an intentional stance was deployed after participants were explicitly instructed to predict behavior, many cases of mental state attribution appear to be driven from bottom-up perceptual features (Arico et al. 2011; Heider and Simmel 1944; Fiala et al. 2011). In other words, the intentional stance is elicited automatically by some external stimuli. Evidence for the bottom-up elicitation of an intentional stance comes from the work of Heider and Simmel (1944), who famously presented “an illusion” in which two smaller shapes were “chased” by a larger one. The shapes were simple geometric figures, sharing few surface features with entities to which we typically attribute mental states. Nonetheless, participants described the short film as depicting a fight between a larger bully triangle and a brave small triangle, a rescue of a small circle by the small triangle, and a furious large triangle smashing up a room in frustration. Describing the shapes in purely mechanical terms would not be incorrect; in fact the patient SM (who suffered from a bilateral amygdala lesion) described the “illusion” in exactly this way, using far fewer affective and social descriptors, and far more movement descriptors than healthy controls (Heberlein and Adolphs 2004). Yet most people cannot help but “see” the shapes as intentional agents—agents who can feel a certain way (e.g., fear, fury) and who can want certain things (e.g., capture, escape).

The largely automatic nature of the attribution of intentional mental states is further supported by its early emergence in development. Hamlin et al. (2007) used shapes (with cartoon eyes) to show 6- to 10-month-old infants a simple story of one shape being helped and hindered respectively by two

other shapes. First, a circle struggles to climb a hill until a square arrives to push him up. Later, the circle struggles to climb the hill again, until a triangle arrives to shove him back down the hill. When infants are later presented with the square and triangle, they prefer to grab the square, presumably due to their understanding of its positive intentions and good character. Dunfield and Kuhlmeier (2010) also demonstrated that by 21 months, infants can understand and draw preferences based on the intentions of adults, even understanding the difference between adults who are accidentally versus intentionally helpful. Infants preferred adults who tried but failed to give a toy to the infant to adults who accidentally provided the desired toy. Thus, our capacity to infer the presence of mental states, and make behavioral predictions based on them, develops early and may reflect a natural source of our social understanding (Waytz et al. 2010).

Adopting an intentional stance—thinking of an entity as having a human-like mind—appears to aid action understanding (or at least perceived action understanding), but this seems to insufficiently capture the full sense of mind we attribute to humans. To humans, we do not only attribute intentions and goals, but also moral rights. Even when we attribute intentions to our chess computer, if we got bored with it then it wouldn't bother us to disassemble it and turn it into a toaster. Mental state attribution might therefore support action understanding on some level as we deal with computers or other inanimate entities, but simply taking an intentional stance or anthropomorphizing an entity does not necessarily imbue it with moral rights. To account for the full extent of our attribution of mind, researchers have begun to converge on the notion that we attribute more than one kind of mind (Gray et al. 2007; Gray et al. 2011; Gray and Wegner 2011a; Jack and Robbins 2012; Knobe and Prinz 2008; Robbins and Jack 2006). In the next section, we will explore how the attribution of moral rights may depend on our ability to empathize and attribute the capacity for pain, pleasure, and emotions.

2 Of two minds: Extending the intentional stance

Philosophers and psychologists have begun to converge on the notion that multiple dimensions of mind attribution compose moral personhood

(Gray et al. 2007; Gray et al. 2011; Gray and Wegner 2011a; Jack and Robbins 2012; Knobe and Prinz 2008; Robbins and Jack 2006). This notion is present even in Dennett's writings (1981b) though not pursued further:

... [When] we declare a man insane we cease treating him as accountable, and we deny him most rights, but still our interactions with him are virtually indistinguishable from normal person interactions unless he is very far gone into madness indeed. In one sense of 'person,' it seems, we continue to treat and view him as a person.

Put plainly, even if we absolve a man of his responsibilities in the case of insanity, we would not excuse ourselves for behaving badly toward him—this person still has rights. This person's pain and pleasure must factor into our moral considerations and how we act toward him.

To account for the observation that moral concern is preserved despite changes in the intentional stance, Robbins and Jack (2006) proposed that we also adopt a *phenomenological stance*. In essence, while the intentional stance involves the attribution of intentions to another person, the phenomenological stance involves the attribution of an emotional experience.

Robbins and Jack (2006) support their dissociation of the intentional and phenomenological stance by drawing attention to a distinction between two sorts of empathy: cognitive empathy and emotional empathy (Davis 1983; Frith 2003; Gonzalez-Liencrez et al. 2013; Shamay-Tsoory et al. 2009; Smith 2006). Robbins and Jack (2006) argue that the intentional stance is supported by cognitive empathy, while the phenomenological stance is supported by emotional empathy.

Emotional empathy describes the shared experience of the affective states of others, which is hypothesized to be a phylogenetically older ability (Decety et al. 2012; Gonzalez-Liencrez et al. 2013; Smith 2006), with likely evolutionary roots in an ability to empathize with kin and offspring (Trivers 1971). While emotional empathy may have originated in empathy for immediate family, it is hypothesized to have been extended to non-kin over time (cf. expanding the moral circle; Singer 1981). Lesions in brain regions such as the inferior frontal gyrus (IFG) and Brodmann area 44 (which has been implicated as part of the mirror neuron system (Rizzolatti 2005)) have produced deficits in emotional empathy specifically (Shamay-Tsoory et al. 2009). Importantly, emotional

empathy is thought to be distinct from the simple ability to vicariously experience emotional arousal, known as emotional contagion (Lorenz 1935). The distinction could be formalized as follows: emotional contagion takes the form of: “You feel X; therefore I feel X,” while emotional empathy takes the form of: “You feel X because Y; therefore I feel X because you feel X” (Gonzalez-Liencre et al. 2013).

Cognitive empathy involves *understanding* the mental states of others, without necessarily *experiencing* the same mental states as the target of the empathy, and could be formalized as follows: “You feel X; therefore I feel Y” (Gonzalez-Liencre et al. 2013). Cognitive empathy is generally considered as theoretically similar to theory of mind, in that both involve an understanding of the representational content of mental states held by others; in addition, measures of cognitive empathy and theory of mind have been shown to correlate (Shamay-Tsoory et al. 2009). Lesions to brain regions that are associated with theory of mind tasks, such as the ventromedial prefrontal cortex (Mitchell et al. 2006), have been associated with deficits in cognitive empathy (Shamay-Tsoory et al. 2009), suggesting that the two processes may also share neural substrates.

The distinction between the intentional stance and the phenomenological stance is supported to some extent by the contrasting deficits in empathy in the cases of autism and psychopathy (Robbins and Jack 2006; Smith 2006). Psychopathy is argued to consist of impaired emotional empathy but preserved cognitive empathy, while autism is argued to consist of impaired cognitive empathy but preserved emotional empathy. Individuals with autism exhibit normal physiological arousal to the distress of others (Blair 1999) and score similar to controls on some emotional empathy tasks (Dziobek et al. 2008); thus, emotional empathy may be preserved. However, the claim that emotional empathy is completely spared in persons with autistic spectrum disorders (ASD) is controversial (Gonzalez-Liencre et al. 2013), due both to some evidence that these persons have difficulty identifying basic emotions in faces (Clark et al. 2008), and to the simple fact that deficits vary so widely in ASD that individual presentations rarely adhere to such a sharp categorical boundary. Nonetheless, several other disorders may fit the profile of impaired cognitive empathy but spared emotional empathy, including frontotemporal lobar degeneration (Rankin et al. 2005), bipolar disorder (Cusi et al. 2010;

Shamay-Tsoory et al. 2009), and borderline personality disorder (Harari et al. 2010; Mier et al. 2013; Minzenberg et al. 2006).

By contrast, in the case of psychopathy, emotional empathy is argued to be impaired, while cognitive empathy remains intact (Robbins and Jack 2006). Psychopaths show impaired autonomic arousal to images depicting distress (Blair et al. 1996) but perform well on tasks measuring theory of mind (Blair et al. 1996); psychopaths even perform well on the “reading the mind in the eyes” task, which requires participants to read off emotions from photographs of eyes (Richell et al. 2003). Other work indicates a similar dissociation in children with psychopathic traits (Jones et al. 2010); these participants show intact theory of mind ability, but reduced concern for other people’s feelings in hypothetical scenarios (e.g. “you forgot your friend’s birthday and made him feel sad”). This deficit in emotional empathy has also been hypothesized to drive psychopaths’ relatively lenient judgments of accidental moral violations (Young et al. 2012). Psychopaths judged accidental harms as morally permissible, presumably due to a failure to be moved emotionally by the harmful outcome, compared to control participants (cf. Cushman 2008). This behavioral profile also stands in contrast to that of ASD participants, who deliver particularly harsh judgments of accidental harms (Moran et al. 2011), due to deficits in encoding agents’ innocent mental states.

Above we provided evidence that observers may have both bottom-up and top-down routes available to the adoption of an intentional stance: bottom-up through low-level perceptual cues (Arico et al. 2011; Fiala et al. 2011; Heider and Simmel 1944), and top-down through explicit motivation to understand behavior (Dennett 1981a; Waytz et al. 2010). Bottom-up processes can certainly drive the adoption of a phenomenological stance; emotional contagion (Lorenz 1935) is foundational to emotional empathy (Gonzalez-Liencre et al. 2013). However, can the phenomenological stance be deployed through top-down processes as well? Some recent evidence hints that this may be the case, and that participants can successfully adopt a phenomenological stance when instructed, even to neutral faces that lack any salient emotional content. Participants’ success in this task is verified through measurable differences in brain activity in areas known to be associated with emotional empathy (de Greck et al. 2012; Nummenmaa et al. 2008). For example, de Greck et al. (2012) instructed participants to view people’s

faces (which were angry or neutral) and either to empathize or to make skin color judgments. These researchers found that empathizing activated bilateral inferior prefrontal cortex, known to be part of the mirror neuron system (Kaplan and Iacoboni 2006). Notably, the inferior prefrontal cortex was preferentially activated when participants actively worked to engage emotional empathy (i.e. adopt a phenomenological stance), compared to when participants made skin color judgments, independent of the emotional content of the face. This result hints at the possibility that emotional empathy can be deployed in a top-down fashion.

What is central to the phenomenological account according to Robbins and Jack (2006) is that witnessing suffering or any aversive phenomenological state should be "primitively morally compelling." In other words, feeling moral concern should flow naturally from recognizing an aversive phenomenological state in others. The work reviewed above on emotional empathy demonstrates that we can easily share and understand the emotions of others, but how are we compelled to care? One potential route is through the "pain matrix," a set of brain regions recruited for both the personal experience of pain, and for witnessing the pain of others (Botvinick et al. 2005; Lamm et al. 2011; Jackson et al. 2006; Singer and Lamm 2009; Singer et al. 2004). According to a recent meta-analysis, the pain matrix includes the anterior medial cingulate cortex (aMCC), posterior anterior cingulate cortex (pACC), and bilateral anterior insula (AI) (Lamm et al. 2011). In one common experimental design (cue-based), participants observe an abstract cue (e.g. a colored light), indicating that either they or a partner will receive a painful electric shock (Lamm et al. 2011). Participants could not see, hear, or touch their partner, and the only indication of the pain their partner would feel was the cue. Regardless of whether the cue indicated the participant or their partner would experience pain, activity in aMCC, pACC, and AI was found in common between the self and other-pain trials (Lamm et al. 2011). Furthermore, neural activity in aMCC, pACC, and AI was also found in common between cue-based, other-pain trials and picture-based studies, in which participants witnessed painful events (such as a car door slamming on someone else's fingers) (Lamm et al. 2011). Cue-based paradigms might model a top-down route to the representation of others' pain, containing no perceptual features typically associated with pain, while picture-based paradigms might model a bottom-up route through their use of painful

imagery. If common neural substrates are involved in the processing of our own pain, and the pain of others, then the morally compelling nature of others' pain may have its roots in the repurposing of neural circuitry that had previously compelled us to be concerned with our own aversive experiences.

If empathy can be driven by bottom-up cues at times, but also can be implemented through top-down processes, then are there any limiting cases where we cannot adopt a phenomenological stance toward an entity? According to work by Knobe and Prinz (2008), adopting a phenomenological stance toward group minds may be one such limiting case. Groups refer to organizations composed of individual group members, such as corporations or political parties, and Knobe and Prinz investigated whether people have the intuition that these groups can be treated as having an analogous mind to an individual. They found that people attribute intentional mental states but not phenomenal mental states toward groups.² For example, the statement "ACME Corp. *believes* that by opening 20 new stores they can increase revenue" is reported to sound natural; however, we might be reluctant to endorse the statement "ACME Corp. feels *depressed* because the expansion fails to generate the expected revenue." This use of mental states to describe groups is not simply a metaphorical use; according to participants it appears to be considered literal (Arico et al. 2011). Knobe and Prinz suggest that people may apply intentional mental states broadly to predict behavior, whereas phenomenal mental states might be linked to moral concern, and their attribution might be constrained by additional features. One of these features might be the possession of a physical body with which to empathize.

However, recent work has conflicted with conclusions of Knobe and Prinz (2008). Huebner et al. (2010) present evidence that the difficulty in attributing phenomenal states to groups might instead stem from a Western cultural bias toward individualism. When students in Hong Kong answered questions similar to those used by Knobe and Prinz (2008), they were more likely to attribute phenomenal mental states to groups as entities than American students. This finding is consistent with other work showing the emphasis of Western cultures on individuals' identities and the emphasis of Eastern cultures on collective identity. Further, Knobe and Prinz (2008) have been criticized for making direct comparisons between groups and individuals (Sytsma and Machery 2009). In particular, groups cannot perform many of the

actions typically associated with phenomenal mental states. If a corporation is depressed, for example, it is not capable of crying, insomnia, irritability, and so forth, all of which are behavioral consequences that afford attributions of depression to an entity (Sytsma and Machery 2009).

Understanding that we can adopt a phenomenological stance, in addition to an intentional stance, goes a long way toward making sense of how we attribute mental states. Above we reviewed work that suggests that these two stances are somewhat dissociable (in clinical cases of autism and psychopathy), and that the phenomenological stance likely underlies moral concern. In the next section, we will review a line of psychological work that arrived at a similar conclusion: that we attribute two dimensions of mind, and that these dimensions have unique roles in the domain of morality.

3 Dimensions of mind perception

While we have focused on how people understand other entities by adopting an intentional stance or a phenomenological stance, a distinct line of psychological work has approached mental state attribution using a dimensional approach (Gray et al. 2007; Gray et al. 2011; Gray et al. 2012a; Gray et al. 2012b; Gray and Wegner 2009, 2010, 2011a, 2011b). On this approach, mental state attributions can be made along two dimensions, Agency and Experience, roughly equivalent to the adoption of an intentional stance and a phenomenological stance. In an initial demonstration (Gray et al. 2007), participants judged the relative mental capacities of babies, robots, dead people, adult humans, god, and so on, on the extent that they possess a variety of mental capacities (e.g. capacity for exercising self-control, capacity for feeling pain, etc.). After analyzing the dimensions for factors that could best explain underlying patterns across the entire set of data, two primary components emerged, accounting for 97 percent of the variance in the observed data. First, accounting for 88 percent of the variance in the data was Experience, which included items relating to the experience of hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy. Second, accounting for an additional 8 percent of the variance in the data was Agency, which included items relating to self-control, morality, memory, emotion recognition, planning, communication, and thought.

Some judgments, such as liking the entity, saving the entity from destruction, making the entity happy, or perceiving the entity as having a soul, were correlated with both Agency and Experience. However, of central importance was the observation that Agency was uniquely related to punishing the entity for causing harm, while Experience was related to an aversion toward harming the entity. This work may then be thought to provide evidence of a link between dimensions of mind perception, and moral rights and responsibilities: moral responsibilities are associated with Agency,³ and moral rights are associated with Experience.

If Agency is associated with moral responsibility and Experience is associated with moral rights then in a typical moral violation involving a perpetrator harming a victim, Agency should be attributed to the perpetrator, who we want to hold responsible for his or her actions, and Experience should be attributed to the victim, whose rights we want to defend. Recent work has made the argument that these associations, combined with the template of a typical moral violation (a perpetrator harming a victim), guide our attribution of mental states (Gray and Wegner 2011a, 2011b; Gray et al. 2012a; Gray et al. 2012b). Based on this template, attributions of Agency and Experience might interact, where attributing more Agency to the perpetrator leads to an increase in Experience attributed to the victim, and vice versa. For example, a harm that is perceived as having been committed intentionally is reported to feel more painful⁴ (Gray 2012; Gray and Wegner 2008). Gray et al. (2012a) and Gray et al. (2012b) broadly refer to this phenomenon as *dyadic completion*, where observers will infer a perpetrator in the presence of a suffering victim and a victim in the presence of a harmful perpetrator. Consistent with this account, Gray and Wegner (2010) found a significant negative correlation between religiosity and a “suffering index” (the inverse of a national health index) across American states. States experiencing the most “suffering” also reported the highest belief in god (controlling for education and median income). While Gray and Wegner (2010) did not explicitly test the converse (inferring a victim in the case of victimless immoral behavior), there is no shortage of intuitive examples: such as believing that drug use or homosexuality is necessarily harmful, even in the absence of concrete evidence of harm.

Further work on the dimensional framework of Agency and Experience has highlighted the dissociation of Agency and Experience in subclinical

populations, based on differences in ASD, psychopathic, and schizophrenic characteristics. Gray et al. (2011) replicated the results of Gray et al. (2007) and in addition collected measures of individual differences on the Autism Spectrum-Quotient Scale (Baron-Cohen et al. 2001), measuring traits related to ASD; the Self-Report Psychopathy Scale (Paulhus et al. 2009), measuring traits related to psychopathy; and the Schizotypy Personality Questionnaire (Raine and Benishay 1995), measuring traits related to schizophrenia. Participants scoring high on the Autism Scale perceived less Agency in entities; participants scoring high on the Psychopathy Scale perceived less Experience in entities; and participants scoring high on the Schizotypy Personality Scale perceived more Agency in entities. These results have been taken as further support for Agency and Experience as largely orthogonal and dissociable dimensions of mind attribution. These results also parallel the deficits discussed above of cognitive empathy in ASD, and emotional empathy in psychopathy.

4 Agency and experience as sources of moral standing

The work we have reviewed so far suggests that Experience is essential to granting moral rights (Gray et al. 2007; Jack and Robbins 2006), while Agency is essential to attributing moral responsibility (Gray et al. 2007). However, Sytma and Machery (2012) have recently suggested that Experience alone cannot completely account for the range of ways in which we attribute moral standing (moral rights). Experience works well to account for the sources of moral standing considered by utilitarian thinkers (Bentham 1781/2000; Singer 1981), but on deontological grounds the moral standing of an entity often depends on its rationality (Kant 1785/2005). In other words, Agency may contribute more to moral standing than previously thought. Utilitarianism maintains that moral decisions should be based on a metric, where the correct moral action is the one that maximizes well-being and minimizes suffering (Bentham 1781/2000). This utilitarian metric relies on victims' capacity for suffering and on observers' capacity to both empathize with others, and desire to prevent their suffering. Kant's deontological morality (1785/2005), on the other hand, emphasizes that the moral standing of victims should be based on the extent to which they are capable of making rational decisions, rather than

simply being driven by passions (emotion) (see Sytsma and Machery (2012) for a thorough review of thinkers in utilitarian and deontological schools of thought). For Kant, the source of moral standing appears to be Agency, rather than Experience. On this basis, Sytsma and Machery (2012) argue that moral standing may derive from two sources: Agency and Experience.

Sytsma and Machery (2012) provide an illustrative example of Agency's role in moral standing in the 1550–51 debate in Valladolid, Spain, over whether the Spanish could rightfully enslave the aboriginal Indians in North America. The debate was between Sepúlveda, who argued for the enslavement of the Indians, and Las Casas, who argued against it. Importantly, the debate centered not on the capacity of the Indians to suffer, but on whether the concept of barbarians could appropriately be applied to them. Sepúlveda argued that the Indians were uncivilized; while Las Casas emphasized that the Indians had sophisticated civilizations and languages, and applied their own rule of law (Sytsma and Machery 2012). Las Casas was successful, and a papal decree was issued, declaring that the Indians were not to be enslaved:

The enemy of the human race, who opposes all good deeds in order to bring men to destruction, beholding and envying (the spreading of the Catholic Faith), invented a means never before heard of, by which he might hinder the preaching of God's word of Salvation to the people: he inspired his satellites who, to please him, have not hesitated to publish abroad that the Indians of the West and the South, and other people of whom We have recent knowledge should be treated as dumb brutes created for our service, pretending that they are incapable of receiving the Catholic Faith.

We, who, though unworthy, exercise on earth the power of our Lord and seek with all our might to bring those sheep of His flock who are outside into the fold committed to our charge, consider, however, that *the Indians are truly men and that they are not only capable of understanding the Catholic Faith but, according to our information, they desire exceedingly to receive it.* (Emphasis added)⁵

The moral status of the Indians in this case did not hinge on whether or not they were capable of suffering, but instead on their rationality. The papacy is broadening the community of the catholic congregation to include the Indians as a result of their capacity and (alleged) desire to share the same set of beliefs.

Experimental evidence provided by Sytsma and Machery (2012) provides further support for the role of Agency in moral standing. Sytsma and Machery (2012) first replicated the previous finding of Experience, but not Agency, driving moral standing (Gray et al. 2007; Gray et al. 2011), using the example of lethal experimentation on monkeys. However, they repeated their initial paradigm in another scenario: would it be acceptable for humans to experiment on the population of a newly discovered alien race? In this case, participants were more opposed to experimentation on alien races with Agency, but the race's capacity for Experience was irrelevant to their moral standing. A follow-up study asked participants about the acceptability of lethal experimentation on a single alien where, in addition to replicating the effect of Agency, a small effect of Experience emerged as well. The results suggest that Experience may play a more prominent role in the moral standing of individuals, rather than the moral standing of groups (cf. Knobe and Prinz 2008).

The potential role of Agency as an additional source of moral standing is promising for interpreting the broad range of moral norms and behaviors. In particular, how moral standing derives from Agency in contrast to Experience could be a particularly fruitful avenue of future research. Witnessing suffering should be "primitively morally compelling" (Robbins and Jack 2006), and we discussed above how this might be so by reviewing the literature surrounding emotional empathy and the vicarious experience of aversive states through the pain matrix. But the mechanism through which Agency grants moral standing remains unclear, and the work reviewed in this chapter seems to suggest that simply attributing Agency is not enough to compel us to grant moral rights to the target (Gray et al. 2007; Gray et al. 2011; Gray et al. 2012a; Gray et al. 2012b). One alternative might be that Agency compels us to take the goals of rational entities seriously (see Kant 1785/2005). If someone were to complain they were being mistreated, and we saw them as rational, then we should either address their complaint, or find a justifiable reason for why we shouldn't have to. In contrast, if a child says being forced to eat broccoli is mistreatment then we should feel comfortable overruling them, as we don't see them as rational to the same extent that we are. Moral standing might then derive from Agency through the reasonable assumption that if an entity were rational, it would want to be treated ethically. For instance, when the Spaniards determined that the Indians were rational, it was not

empathy that was responsible for the Spaniards' restraint, but respect for the expressed desire of the Indians *to not be enslaved*. A counterexample from Douglas Adams's *Restaurant at the End of the Universe* makes the role of the expressed goals of a rational entity even more apparent by providing an Agentic creature that advocates *against* preserving its life. While traveling through space, Arthur Dent, a human, and his fellow travelers arrive at the restaurant at the end of the universe, where they meet the main course: a cow that wants to be eaten (Adams 1980, Chapter 17). Arthur is disgusted by the cow's recommendation of which body parts to consume, and asks to have a green salad.

'A green salad?' said the animal, rolling his eyes disapprovingly at Arthur.

'Are you going to tell me,' said Arthur, 'that I shouldn't have a green salad?'

'Well,' said the animal, 'I know many vegetables that are very clear on that point. Which is why it was eventually decided to cut through the whole tangled problem and breed an animal that actually wanted to be eaten and was capable of saying so clearly and distinctly. And here I am.'

Clearly the cow's invitation doesn't solve the dietary dilemma for Arthur and indeed introduces a new dilemma between Arthur's aversion to killing the cow (the cow's Experience), and the cow's insistence on being killed and eaten (the cow's Agency). This example reveals that the cow's Agency can influence its moral standing, but factors such as Experience will continue to have influence, even against the cow's expressed desire to be killed.

5 Conclusion

In this chapter we've reviewed the theory behind the intentional stance, and shown that it is central to how we make predictions about complex behaviors. Despite its importance, the intentional stance (or Agency) in isolation cannot completely account for our attribution of moral standing toward entities. Recognizing that we also adopt a phenomenological stance (attributing Experience) toward entities, and that this happens largely independently of our attribution of Agency, provides a more complete picture of our understanding

of mental states. Ultimately, however, moral standing may not be the exclusive domain of either Agency or Experience (Sytsma and Machery 2012), and by recognizing the multiple sources of our moral standing we might come closer to capturing the entirety of our moral universe.

Notes

- 1 Although actually solving the problem requires us to adopt a *physical* stance toward the clock, Waytz et al. (2010) found that people are even more likely to adopt an *intentional* stance in anthropomorphizing the malfunctioning clock. Thus, when an entity appears to be beyond easy understanding, we may resort to attributing mental states to that entity (e.g., my clock *forgot* to wake me up because it is *out to get me*).
- 2 More specifically, Knobe and Prinz tested folk intuitions about phenomenal consciousness: the second-order property that there “is something to be like” in states such as seeing red, hearing a C# musical note, etc. (Sytsma and Machery 2010). In particular, Knobe and Prinz wanted to test whether nonphilosophers understand this philosophical concept. Sytsma and Machery (2010) have argued that Knobe and Prinz conflate the folk understanding of this concept with the folk understanding of subjective experience (i.e. the phenomenal stance). We leave aside this debate surrounding phenomenal consciousness and simply present the work by Knobe and Prinz as an illustration of the relative independence with which the intentional and phenomenological mental states can be attributed.
- 3 One might notice that Agency involves the attribution of moral responsibility, while adopting an intentional stance does not necessarily implicate the target as morally responsible. The boundaries of where the capacity for intentional action gives way to moral responsibility have, to our knowledge, not been well explored. However, based on the history of moral philosophy exploring the moral consequences of being a rational, self-determining actor (Sytsma and Machery 2012), we suggest that the boundary has to do with seeing a target as capable of rationally forming its own goals. Remember that adopting an intentional stance allows us to assume a target will act rationally to achieve its goals, but that does not necessarily mean it can choose them. For example, the chess computer certainly never chose to desire to win at chess.

- 4 The claim that the Agency of the perpetrator affects the subjective experience of pain in the victim is beyond the focus of this chapter. To reiterate, we are specifically concerned with the attribution of mental states to others as opposed to the subjective experience of one's own mental states.
- 5 <http://www.papalencyclicals.net/Paul03/p3subli.htm>

References

- Adams, D. (1980/2010), *The Restaurant at the End of the Universe*. London: Pan Macmillan Limited.
- Arico, A., Fiala, B., Goldberg, R. F., and Nichols, S. (2011), "The folk psychology of consciousness." *Mind & Language*, 26(3), 327–52.
- Baron-Cohen, S. (2001), "Theory of mind in normal development and autism." *Prisme*, 34, 174–83.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001), "The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians." *Journal of Autism and Developmental Disorders*, 31(1), 5–17.
- Bentham, J. (1781/2000), *An Introduction to the Principles of Morals and Legislation*. Kitchener: Batoche Books.
- Blair, R. J. R. (1999), "Psychophysiological responsiveness to the distress of others in children with autism." *Personality and Individual Differences*, 26, 485.
- Blair, R. J. R., Sellars, C., Strickland, I., Clark, F., Williams, A., Smith, M., and Jones, L. (1996), "Theory of mind in the psychopath." *Journal of Forensic Psychiatry*, 7, 15–25.
- Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., and Prkachin, K. M. (2005), "Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain." *Neuroimage*, 25, 312–19.
- Clark, T. F., Winkielman, P., and McIntosh, D. N. (2008), "Autism and the extraction of emotion from briefly presented facial expressions: Stumbling at the first step of empathy." *Emotion*, 8(6), 803–9.
- Cushman, F. (2008), "Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment." *Cognition*, 108(2), 353–80.
- Cusi, A., Macqueen, G. M., and McKinnon, M. C. (2010), "Altered self-report of empathic responding in patients with bipolar disorder." *Psychiatry Research*, 178(2), 354–8.
- Davis, M. H. (1983), "Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of Personality and Social Psychology*, 44(1), 113–26.

- De Greck, M., Wang, G., Yang, X., Wang, X., Northoff, G., and Han, S. (2012), “Neural substrates underlying intentional empathy.” *Social Cognitive and Affective Neuroscience*, 7(2), 135–44.
- Decety, J., Norman, G. J., Berntson, G. G., and Cacioppo, J. T. (2012), “A neurobehavioral evolutionary perspective on the mechanisms underlying empathy.” *Progress in Neurobiology*, 98(1), 38–48.
- Dennett, D. (1981a), “Intentional systems,” in D. Dennett (ed.), *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: MIT Press, pp. 3–22.
- (1981b), “Conditions of personhood,” in D. Dennett (ed.), *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: MIT Press, pp. 267–85.
- (1987), *The Intentional Stance*. Cambridge: MIT Press.
- Dunfield, K. A. and Kuhlmeier, V. A. (2010), “Intention-mediated selective helping in infancy.” *Psychological Science*, 21(4), 523–7.
- Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., and Convit, A. (2008), “Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET).” *Journal of Autism and Developmental Disorders*, 38, 464–73.
- Fiala, B., Arico, A., and Nichols, S. (2011), “On the psychological origins of dualism: Dual-process cognition and the explanatory gap,” in E. Slingerland and M. Collard (eds), *Creating Consilience: Integrating the Sciences and the Humanities*. Oxford: Oxford University Press, pp. 88–110.
- Frith, U. (2003), *Autism: Explaining the Enigma* (2nd edition). Oxford: Blackwell.
- Gonzalez-Liencre, C., Shamay-Tsoory, S. G., and Brüne, M. (2013), “Towards a neuroscience of empathy: Ontogeny, phylogeny, brain mechanisms, context and psychopathology.” *Neuroscience and Biobehavioural Reviews*, 37(8), 1537–48.
- Gray, K. (2012), “The power of good intentions: Perceived benevolence soothes pain, increases pleasure, and improves taste.” *Social Psychological and Personality Science*, 3(5), 639–45.
- Gray, K. and Wegner, D. M. (2008), “The sting of intentional pain.” *Psychological Science*, 19(12), 1260–62.
- (2009), “Moral typecasting: Divergent perceptions of moral agents and moral patients.” *Journal of Personality and Social Psychology*, 96(3), 505–20.
- (2010), “Blaming god for our pain: Human suffering and the divine mind.” *Personality and Social Psychology Review*, 14(1), 7–16.
- (2011a), “Morality takes two: Dyadic morality and mind perception,” in M. Mikulincer and P. Shaver (eds), *The Social Psychology of Morality: Exploring the Causes of Good and Evil*. Washington, DC: American Psychological Association, pp. 109–27.

- (2011b), “To escape blame, don’t be a hero—be a victim.” *Journal of Experimental Social Psychology*, 47(2), 516–19.
- Gray, H. M., Gray, K., and Wegner, D. M. (2007), “Dimensions of mind perception.” *Science*, 315, 619.
- Gray, K., Waytz, A., and Young, L. (2012a), “The moral dyad: A fundamental template unifying moral judgment.” *Psychological Inquiry*, 23(2), 206–15.
- Gray, K., Young, L., and Waytz, A. (2012b), “Mind perception is the essence of morality.” *Psychological Inquiry*, 23(2), 101–24.
- Gray, K., Jenkins, A. C., Heberlein, A. S., and Wegner, D. M. (2011), “Distortions of mind perception in psychopathology.” *Proceedings of the National Academy of Sciences*, 108, 477–9.
- Hamlin, J. K., Wynn, K., and Bloom, P. (2007), “Social evaluation by preverbal infants.” *Nature*, 450, 557–9.
- Harari, H., Shamay-Tsoory, S. G., Ravid, M., and Levkovitz, Y. (2010), “Double dissociation between cognitive and affective empathy in borderline personality disorder.” *Psychiatry Research*, 175(3), 277–9.
- Heberlein, A. S. and Adolphs, R. (2004), “Impaired spontaneous anthropomorphizing despite intact perception and social knowledge.” *Proceedings of the National Academy of Sciences*, 101(19), 7487–91.
- Heider, F. and Simmel M. (1944), “An experimental study of apparent behavior.” *American Journal of Psychology*, 57, 243–59.
- Huebner, B., Bruno, M., and Sarkissian, H. (2010), “What does the nation of China think about phenomenal states?” *Review of Philosophy and Psychology*, 1(2), 225–43.
- Jack, A. I. and Robbins, P. (2012), “The phenomenal stance revisited.” *Review of Philosophy and Psychology*, 3(3), 383–403.
- Jackson, P. L., Rainville, P., and Decety, J. (2006), “To what extent do we share the pain of others? Insight from the neural bases of pain empathy.” *Pain*, 125, 5–9.
- Jones, A. P., Happé, F. G. E., Gilbert, F., Burnett, S., and Viding, E. (2010), “Feeling, caring, knowing: Different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder.” *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 51(11), 1188–97.
- Kant, I. (1785/2005), *Groundwork for the Metaphysics of Morals*. Toronto: Broadview Press.
- Kaplan, J. T. and Iacoboni, M. (2006), “Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy.” *Social Neuroscience*, 1(3–4), 175–83.
- Knobe, J. and Prinz, J. (2008), “Intuitions about consciousness: Experimental studies.” *Phenomenology and the Cognitive Sciences*, 7, 67–83.

- Lamm, C., Decety, J., and Singer, T. (2011), “Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain.” *NeuroImage*, 54(3), 2492–502.
- Lorenz, K. (1935), “Der Kumpan in der Umwelt des Vogels.” *Journal für Ornithologie*, 83(2), 137–213.
- Mier, D., Lis, S., Esslinger, C., Sauer, C., Hagenhoff, M., Ulferts, J., Gallhofer, B., and Kirsch, P. (2013), “Neuronal correlates of social cognition in borderline personality disorder.” *Social Cognitive and Affective Neuroscience*, 8(5), 531–7.
- Minzenberg, M. J., Poole, J. H., and Vinogradov, S. (2006), “Social-emotion recognition in borderline personality disorder.” *Comprehensive Psychiatry*, 47(6), 468–74.
- Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (2006), “Dissociable medial prefrontal contributions to judgments of similar and dissimilar others.” *Neuron*, 50(4), 655–63.
- Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O’Young, D., Mavros, P. L., and Gabrieli, J. D. (2011), “Impaired theory of mind for moral judgment in high-functioning autism.” *Proceedings of the National Academy of Sciences*, 108(7), 2688–92.
- Nummenmaa, L., Hirvonen, J., Parkkola, R., and Hietanen, J. K. (2008), “Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy.” *NeuroImage*, 43(3), 571–80.
- Onishi, K. H. and Baillargeon, R. (2005), “Do 15-month-old infants understand false beliefs?” *Science*, 308, 255–8.
- Paulhus, D. L., Hemphill, J., and Hare, R. (2009), *Manual of the Self-Report Psychopathy Scale (SRP-III)*. Toronto: Multi-Health Systems.
- Raine, A. and Benishay, D. (1995), “The SPQ-B: A brief screening instrument for schizotypal personality disorder.” *Journal of Personality Disorders*, 9, 346–55.
- Rankin, K. P., Kramer, J. H., and Miller, B. L. (2005), “Patterns of cognitive and emotional empathy in frontotemporal lobar degeneration.” *Cognitive and Behavioral Neurology*, 18(1), 28–36.
- Richell, R. A., Mitchell, D. G. V., Newman, C., Leonard, A., Baron-Cohen, S., and Blair, R. J. R. (2003), “Theory of mind and psychopathy: Can psychopathic individuals read the ‘language of the eyes’?” *Neuropsychologia*, 41, 523–6.
- Rizzolatti, G. (2005), “The mirror neuron system and its function in humans.” *Anatomy and Embryology*, 210(5–6), 419–21.
- Robbins, P. and Jack, A. I. (2006), “The phenomenal stance.” *Philosophical Studies*, 127, 59–85.
- Saxe, R. (2009), “The happiness of the fish: Evidence for a common theory of one’s own and others’ actions,” in K. D. Markman, W. M. P. Klien, and J. A. Suhr (eds), *The Handbook of Imagination and Mental Simulation*. New York: Psychology Press, pp. 257–66.

- Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009), "Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions." *Brain*, 132(3), 617–27.
- Shamay-Tsoory, S., Harari, H., Szepsenwol, O., and Levkovitz, Y. (2009), "Evidence of impaired cognitive empathy in disorder." *Journal of Neuropsychiatry and Clinical Neuroscience*, 21(1), 59–67.
- Silver, N. (2012), *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't*. New York, NY: The Penguin Press.
- Singer, P. (1981), *The Expanding Circle: Ethics and Sociobiology*. New York, NY: Farrar, Straus & Giroux.
- Singer, T. and Lamm, C. (2009), "The social neuroscience of empathy." *Annals of the New York Academy of Sciences*, 1156, 81–96.
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004), "Empathy for pain involves the affective but not sensory components of pain." *Science*, 303, 1157–62.
- Smith, A. (2006), "Cognitive empathy and emotional empathy in human behavior and evolution." *The Psychological Record*, 56, 3–21.
- Sytsma, J. and Machery, E. (2009), "How to study folk intuitions about phenomenal consciousness." *Philosophical Psychology*, 22(1), 21–35.
- (2010), "Two conceptions of subjective experience." *Philosophical Studies*, 151(2), 299–327.
- (2012), "The two sources of moral standing." *Review of Philosophy and Psychology*, 3(3), 303–24.
- Trivers, R. (1971), "The evolution of reciprocal altruism." *The Quarterly Review of Biology*, 46(1), 35–57.
- Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J., and Cacioppo, J. T. (2010), "Making sense by making sentient: Effectance motivation increases anthropomorphism." *Journal of Personality and Social Psychology*, 99(8), 410–35.
- Young, L., Koenigs, M., Kruepke, M., and Newman, J. P. (2012), "Psychopathy increases perceived moral permissibility of accidents." *Journal of Abnormal Psychology*, 121(3), 659–67.